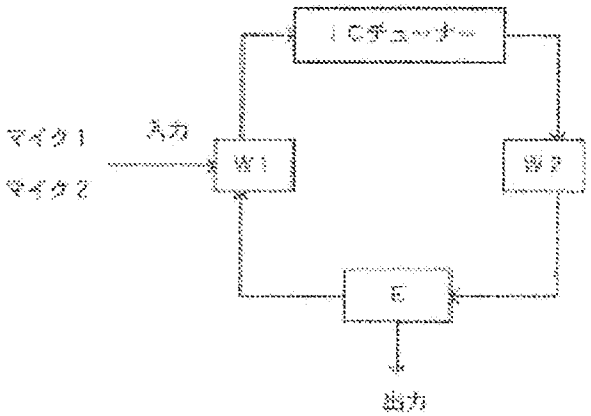
METHOD AND DEVICE FOR VOICE SEPARATION OF COMPOUND VOICE DATA, METHOD AND DEVICE FOR SPECIFYING SPEAKER, COMPUTER PROGRAM, AND RECORDING MEDIUM

Publication number: JP2003005790 (A)
Publication date: 2003-01-08
Inventor(s): YAMAMOTO TAKAYOSHI +
Applicant(s): YAMAMOTO TAKAYOSHI; URATA TAKAYUKI +
Classification:
- International: ***G10L11/00; G10L15/00; G10L15/02; G10L15/20; G10L17/00; G10L21/02;*** **G10L11/00; G10L15/00; G10L17/00; G10L21/00;** (IPC1-7): G10L11/00; G10L15/02; G10L15/20; G10L17/00; G10L21/02
- European:
Application number: JP20010191289 20010625
Priority number(s): JP20010191289 20010625

Also published as:
JP3364487 (B2)

Abstract of **JP 2003005790 (A)**

PROBLEM TO BE SOLVED: To provide a method and a device for separating compound voice data where voice data of several speakers mixedly exist into the voice of every speaker and to provide a method and a device for accurately and quickly specifying the speaker of each separated voice data. SOLUTION: The method for separating compound voice data where voice data of several speakers mixedly exist into the voice data of every speaker has a step (1) where correlation elimination processing is performed to eliminate correlation between the compound voice data and a step (2) where independent component separation processing is performed to separate data subjected to the correlation elimination processing into independent components.



(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2003－5790
（P2003－5790A）
(43)公開日　平成15年１月８日(2003.1.8)

(51)Int.Cl.'	識別記号	ＦＩ		テーマコード'（参考）
Ｇ１０Ｌ　17/00		Ｇ１０Ｌ　3/00	５４５Ａ	５Ｄ０１５
11/00		9/00	Ａ	
15/02		9/02	Ａ	
15/20		9/06	Ａ	
21/02		9/02	３０１Ａ	

審査請求　有　請求項の数21　ＯＬ　（全 32 頁）　最終頁に続く

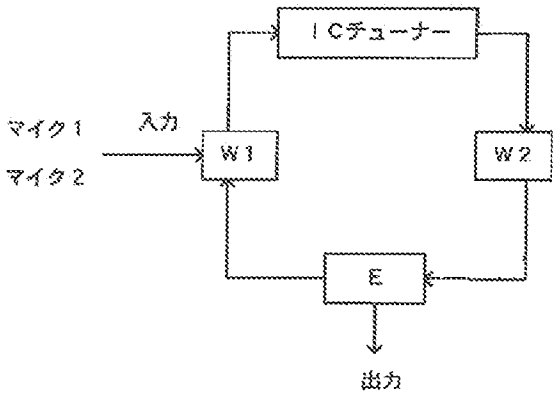
(21)出願番号　特願2001－191289(P2001－191289)
(22)出願日　平成13年６月25日(2001.6.25)

(71)出願人　500262511
山本　隆義
広島市西区己斐上２丁目54－20
(71)出願人　501253822
浦田　隆之
広島県広島市安佐南区高取南１丁目３－５－８
(72)発明者　山本　隆義
広島県広島市西区己斐上２丁目54－20
(74)代理人　100071283
弁理士　一色　健輔　（外３名）

Ｆターム(参考)　5D015 AA03 CC04 CC05

(54)【発明の名称】　複合音声データの音声分離方法、発言者特定方法、複合音声データの音声分離装置、発言者特定装置、コンピュータプログラム、及び、記録媒体

(57)【要約】
【課題】　複数の発言者の音声データが混在する混在音声データを、発言者毎の音声に分離する方法及び装置、さらに分離された各音声データの発言者を特定することを、正確にかつ高速に行うことができる方法及び装置を提供する。
【解決手段】　複数発言者の音声データが混在している混在音声データを、発言者毎の音声データに分離する音声データ分離方法において、（１）前記混在音声データを互いに無相関化するための無相関化処理を行うステップと、（２）前記無相関化処理の行われたデータを独立成分に分離するための独立成分分離処理を行うステップとを有する。



1

【特許請求の範囲】

【請求項１】　複数発言者の音声データが混在している混在音声データを、発言者毎の音声データに分離する音声データ分離方法において、（１）前記混在音声データを互いに無相関化するための無相関化処理を行うステップと、（２）前記無相関化処理の行われたデータを独立成分に分離するための独立成分分離処理を行うステップと、を有することを特徴とする音声分離方法。

【請求項２】　請求項１に記載の音声分離方法において、

前記独立成分分離の行われたデータの分離性が不十分な場合には、分離性が十分になるまで、前記独立成分分離処理の行われたデータについて、前記無相関化処理及び前記独立成分分離処理を繰り返し行うことを特徴とする音声分離方法。

【請求項３】　請求項１又は請求項２に記載の音声分離方法において、

前記独立成分分離処理として、非ガウス性のデータを独立成分に分離するための非ガウス性独立成分分離処理と、非定常性のデータを独立成分に分離するための非定常性独立成分分離処理と、有色性のデータを独立成分に分離するための有色性独立成分分離処理とを準備し、データの性質により、前記非ガウス性独立成分分離処理、前記非定常性独立成分分離処理、及び、前記有色性独立成分分離処理のうちのいずれかの処理を行うことを特徴とする音声分離方法。

【請求項４】　請求項３に記載の音声分離方法において、

最初に行われる独立成分分離処理は、非ガウス性のデータを独立成分に分離するための非ガウス性独立成分分離処理であることを特徴とする音声分離方法。

【請求項５】　請求項１乃至請求項４に記載の音声分離方法において、

前記無相関化処理は、少なくとも主成分分析及び因子分析を行うことを特徴とする音声分離方法。

【請求項６】　複数発言者の音声データが混在している混在音声データを、発言者毎の音声データに分離し、該発言者毎の音声データにつき発言者を特定する発言者特定方法において、（１）請求項１乃至請求項５のいずれかに記載の音声分離方法により、複数発言者の音声データが混在している混在音声データを、発言者毎の音声データに分離するステップと、（２）発言者毎に該発言者を特定するための特定パラメータを準備するステップと、（３）分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して、発言者を特定するステップと、を有することを特徴とする発言者特定方法。

【請求項７】　請求項６に記載の発言者特定方法において、

前記特定パラメータは、発言者が母音を発音した際のホルマント周波数であり、

分離された前記発言者毎の音声データにつき、ホルマント周波数を求め、求められたホルマント周波数に関して、前記特定パラメータとしてのホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定方法。

【請求項８】　請求項７に記載の発言者特定方法において、

前記特定パラメータは、発言者が母音を発音した際の第１ホルマント周波数及び第２ホルマント周波数であり、

分離された前記発言者毎の音声データにつき、第１ホルマント周波数及び第２ホルマント周波数を求め、求められた第１ホルマント周波数及び第２ホルマント周波数に関して、前記特定パラメータとしての第１ホルマント周波数及び第２ホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定方法。

【請求項９】　請求項６乃至請求項８のいずれかに記載の発言者特定方法において、

分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して発言者を特定するステップにて発言者を特定できなかった場合には、

該音声データから複数の時点のホルマント周波数を求め、求められた複数時点のホルマント周波数に関して、前記特定パラメータとしての複数時点のホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定方法。

【請求項１０】　請求項６乃至請求項９のいずれかに記載の発言者特定方法において、

分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して発言者を特定するステップにて発言者を特定できなかった場合には、

該音声データから有声音データを分離し、該有声音データにつき、ホルマント周波数を求め、求められたホルマント周波数に関して、前記特定パラメータとしてのホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定方法。

【請求項１１】　請求項１０に記載の発言者特定方法において、

分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して発言者を特定するステップにて発言者を特定できなかった場合には、

該音声データから有声音データを分離し、該有声音データにつき、第１ホルマント周波数及び第２ホルマント周波数を求め、求められた第１ホルマント周波数及び第２ホルマント周波数に関して、前記特定パラメータとしての第１ホルマント周波数及び第２ホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定方法。

【請求項１２】　請求項１０または請求項１１に記載の発言者特定方法において、

分離された前記発言者毎の音声データにつき，前記特定パラメータを参照して発言者を特定するステップにて発言者を特定できなかった場合には，
該有声音データにつき、複数の時点のホルマント周波数を求め、求められた複数時点のホルマント周波数に関して、前記特定パラメータとしての複数時点のホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定方法。

【請求項１３】　請求項１０又は請求項１１に記載の発言者特定方法において、
前記音声データから前記有声音データを分離する際に、該音声データに対して独立成分に分離するための独立成分分離処理が行われることを特徴とする発言者特定方法。

【請求項１４】　複数発言者の音声データが混在している混在音声データから，議事録を作成する議事録作成方法において、
請求項6乃至請求項１３のいずれかに記載の発言者特定方法により、分離された前記発言者毎の音声データにつき，発言者を特定するステップと、
特定された発言者と、該発言者の発言とを対応付けて記録媒体に出力することにより、議事録を作成するステップと、を有することを特徴とする議事録作成方法。

【請求項１５】　複数発言者の音声データが混在している混在音声データを、発言者毎の音声データに分離する音声データ分離装置において，
前記混在音声データを互いに無相関化するために無相関化処理を行い、
前記無相関化処理の行われたデータを独立成分に分離するために独立成分分離処理を行うことを特徴とする音声分離装置。

【請求項１６】　請求項１５に記載の音声分離装置において，
前記独立成分分離の行われたデータの分離性が不十分な場合には，分離性が十分になるまで，前記独立成分分離処理の行われたデータについて、前記無相関化処理及び前記独立成分分離処理を繰り返し行うことを特徴とする音声分離装置。

【請求項１７】　請求項１５又は請求項１６に記載の音声分離装置において、
データの性質により、前記独立成分分離処理として、非ガウス性のデータを独立成分に分離するための非ガウス性独立成分分離処理，非定常性のデータを独立成分に分離するための非定常性独立成分分離処理，有色性のデータを独立成分に分離するための有色性独立成分分離処理、のうちのいずれかの処理を行うことを特徴とする音声分離装置。

【請求項１８】　請求項１７に記載の音声分離装置において、
最初に行われる独立成分分離処理は、非ガウス性のデータを独立成分に分離するための非ガウス性独立成分分離処理であることを特徴とする音声分離装置。

【請求項１９】　請求項１５乃至請求項１８に記載の音声分離装置において、
前記無相関化処理は、少なくとも主成分分析及び因子分析を行うことを特徴とする音声分離装置。

【請求項２０】　複数発言者の音声データが混在している混在音声データを、発言者毎の音声データに分離し、該発言者毎の音声データにつき発言者を特定する発言者特定装置において、
請求項１５乃至請求項１９のいずれかに記載の音声分離装置により、複数発言者の音声データが混在している混在音声データを、発言者毎の音声データに分離し、
分離された前記発言者毎の音声データにつき、発言者毎に該発言者を特定するための特定パラメータを参照して発言者を特定することを特徴とする発言者特定装置。

【請求項２１】　請求項２０に記載の発言者特定装置において、
前記特定パラメータは、発言者が母音を発音した際のホルマント周波数であり、
分離された前記発言者毎の音声データにつき、ホルマント周波数を求め，求められたホルマント周波数に関して，前記特定パラメータとしてのホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定装置。

【請求項２２】　請求項２１に記載の発言者特定装置において、
前記特定パラメータは、発言者が母音を発音した際の第１ホルマント周波数及び第２ホルマント周波数であり、
分離された前記発言者毎の音声データにつき、第１ホルマント周波数及び第２ホルマント周波数を求め、求められた第１ホルマント周波数及び第２ホルマント周波数に関して、前記特定パラメータとしての第１ホルマント周波数及び第２ホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定装置。

【請求項２３】　請求項２０乃至請求項２２のいずれかに記載の発言者特定装置において、
分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して発言者を特定できなかった場合には，
該音声データから複数の時点のホルマント周波数を求め、求められた複数時点のホルマント周波数に関して、前記特定パラメータとしての複数時点のホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定装置。

【請求項２４】　請求項２０乃至請求項２３のいずれかに記載の発言者特定装置において、
分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して発言者を特定できなかった場合には，

該音声データから有声音データを分離し、該有声音データにつき、ホルマント周波数を求め、求められたホルマント周波数に関して、前記特定パラメータとしてのホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定装置。

【請求項２５】　請求項２４に記載の発言者特定装置において、

分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して発言者を特定できなかった場合には、

該音声データから有声音データを分離し、該有声音データにつき、第１ホルマント周波数及び第２ホルマント周波数を求め、求められた第１ホルマント周波数及び第２ホルマント周波数に関して、前記特定パラメータとしての第１ホルマント周波数及び第２ホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定装置。

【請求項２６】　請求項２４または請求項２５に記載の発言者特定装置において、

分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して発言者を特定できなかった場合には、

該有声音データにつき、複数の時点のホルマント周波数を求め、求められた複数時点のホルマント周波数に関して、前記特定パラメータとしての複数時点のホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定装置。

【請求項２７】　請求項２４又は請求項２５に記載の発言者特定装置において、

前記音声データから前記有声音データを分離する際に、該音声データに対して独立成分に分離するための独立成分分離処理が行われることを特徴とする発言者特定装置。

【請求項２８】　複数発言者の音声データが混在している混在音声データから、議事録を作成する議事録作成装置において、

請求項２０乃至請求項２７のいずれかに記載の発言者特定装置により、分離された前記発言者毎の音声データにつき、発言者を特定し、

特定された発言者と、該発言者の発言とを対応付けて記録媒体に出力することにより、議事録を作成することを特徴とする議事録作成装置。

【請求項２９】　請求項１乃至請求項５のいずれかに記載の音声分離方法を音声分離装置に実行させるためのコンピュータプログラム。

【請求項３０】　請求項６乃至請求項１３のいずれかに記載の発言者特定方法を発言者特定装置に実行させるためのコンピュータプログラム。

【請求項３１】　請求項２９又は請求項３０に記載のコンピュータプログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、複数発言者の複合音声データの音声を分離する方法、分離したそれぞれの音声データの発言者を特定する方法、複数発言者の複合音声データの音声を分離する装置、分離したそれぞれの音声データの発言者を特定する装置、コンピュータプログラム、及び、記録媒体に関する。

【０００２】

【従来の技術】複数の発言者の音声が混合されて記録されている、音声記録媒体中の複合音声データを、発言者毎に正確に分離する技術が切望されている。具体的には、複合音声データを、音声の入力と同時進行的に発言者毎に分離し特定することで、会議の議事録作成を自動的に行うことのできるような技術が切望されている。

【０００３】従来、長時間にわたる会議の議事録を作成するには、各種の音声記録機器に記録した会議の音声データを、議事録作成担当者が全て聞きなおし、要約するなどして議事録を作成していた。この作業は、音声記録機器の再生と一時停止を何度も繰り返しつつ行う必要があり、手間と時間がかかる。

【０００４】また、もう１つの問題は、発言者の特定が困難であることである。本人が会議に出席した担当者ならまだしも、そうでない担当者が議事録を作成するのは、どの音声がどの発言者によるものなのかを判断するのは非常に困難なことであった。

【０００５】従来、混合音声データからの音声分離、発言者特定に関する技術は幾つか存在してはいるが、１本のマイクに複数人の音声やノイズが混合されて入力される場合でも分離、特定を正確に行い、さらに、複合音声の入力と同時進行的に高速な分離・特定処理を行うことは、時間的に連続な音素データのセグメンテーション、及び調音結合の点で非常に難しい課題であった。

【０００６】特開２００１-２７８９５には、複数の信号源からの音響信号を分離し、所望の信号を合成出力するための信号分離方法が記載されている。この発明は、解析対象の混合音声・音響信号に対し時間・周波数解析を行い、周波数成分の倍音構成を得る。倍音周波数成分のうち、立上がり時間及び立下り時間の少なくとも一方が共通であるか否かで、同一信号源からの周波数成分であるかどうかを同定する。その周波数成分を抽出・再構成することにより、単一信号源からの信号を分離する。

【０００７】この発明は、混合された信号の相関性や独立性といった事項を考慮していないので、同じ周波数帯域に属する混合信号、あるいは同時間帯に存在する混合信号を分離することは困難である。

【０００８】また、特開２０００－９７７５８に記載された音源信号推定装置では、複数の音響信号がそれぞれ混在して複数のチャンネルを介して入力されたときに、

各音源信号が混合係数ベクトルと内積演算されて他の音源信号に加算される混合過程モデルに基づき、混合係数ベクトルに対応する分離係数ベクトルを逐次修正しながら求め、この分離係数ベクトルを用いて音源信号の推定、分離を行う（ＩＣＡの手法）にあたり、分離係数ベクトルの逐次修正に用いる修正ベクトルを正規化する音声信号とそれ以外の信号が相互に混在している信号からそれぞれの信号を推定し、分離するに際し、それぞれの信号パワー変動による推定、分離への影響を軽減することができ、さらに、収束係数を大きくすることができることから安定かつ高速の信号分離が可能となる、とされている。

【０００９】この発明は、独立成分解析（ＩＣＡ）をベースとして分離係数ベクトルを逐次修正しながら行うので、信号パワーの変動影響を軽減でき、高速分離を実現するものであるが、様々な信号源からの音源信号はお互いに独立性を保持しているとは限らない。一般に、たとえ独立した信号源からの音源信号であっても混合されると相関性を有してしまっていることが多いが、その点が考慮されていない。

【００１０】また、特開平９－２５８７８８には、基本周波数の近接した混合音声を適切に区別分離し、音源の数に制限されず、高品質の分離音声を得ることを目的とした音声分離方法および装置が記載されている。この発明では、入力音響信号中に含まれる音声信号の有声音部分と無声音部分の内の有声音部分は有声音の音源方向の情報を加味しながら個別に抽出し、抽出された有声音部分を複数の有声音に分化して有声音の群として抽出し、音声信号の無声音部分は入力音響信号から有声音部分を減算して抽出した残差から各有声音の群の無声音に相当する音響信号の成分として抽出し、各別に抽出された有声音の群に無声音を補充して音声信号を抽出することによって上記目的を実現する。

【００１１】この発明は、音源方位の情報を抽出する音源定位部を有しているが、同じ方向から異なる音声が発せられた場合は分離が困難となる。また、複数の発言者が同じ母音、あるいは有声音を発したときはそれらの分離が困難であると思われる。

【００１２】

【発明が解決しようとする課題】以上のような従来技術が有する種々の問題点を解決すべく、本発明は、複数の発言者の音声データが混在する混在音声データを、発言者毎の音声に分離する方法及び装置、さらに分離された各音声データの発言者を特定することを、正確にかつ高速に行うことができる方法及び装置の提供を主たる目的とする。

【００１３】

【課題を解決するための手段】上記の課題を解決するために、本出願に係る第１の発明は、複数発言者の音声データが混在している混在音声データを、発言者毎の音声データに分離する音声データ分離方法において、（１）前記混在音声データを互いに無相関化するための無相関化処理を行うステップと、（２）前記無相関化処理の行われたデータを独立成分に分離するための独立成分分離処理を行うステップとを有することを特徴とする音声分離方法である。このような第１の発明によれば、入力される混在音声データ（生データ）に含まれる各音声データの相関性、及び独立性の両性質をともに考慮し、複数の音声データや混入する雑音などの有する相関性や独立性が、時間的・空間的に変動する場合でも、発言者毎の音声に正確に分離することができる。

【００１４】また、本出願に係る第２の発明は、第１の発明である音声分離方法において、前記独立成分分離の行われたデータの分離性が不十分な場合には、分離性が十分になるまで、前記独立成分分離処理の行われたデータについて、前記無相関化処理及び前記独立成分分離処理を繰り返し行うことを特徴とする音声分離方法である。このような第２の発明によれば、混在音声データを音源毎の音声データに充分に分離させることができる。

【００１５】また、本出願に係る第３の発明は、第１又は第２の発明である音声分離方法において、前記独立成分分離処理として、非ガウス性のデータを独立成分に分離するための非ガウス性独立成分分離処理と、非定常性のデータを独立成分に分離するための非定常性独立成分分離処理と、有色性のデータを独立成分に分離するための有色性独立成分分離処理とを準備し、データの性質により、前記非ガウス性独立成分分離処理、前記非定常性独立成分分離処理、及び、前記有色性独立成分分離処理のうちのいずれかの処理を行うことを特徴とする音声分離方法である。このような第３の発明によれば、無相関化処理の行われたデータの性質に応じて最適な独立成分分離処理を行うことができるから、混在音声データを音源毎の音声データにより効果的に分離させることができる。

【００１６】また、本出願に係る第４の発明は、第３の発明である音声分離方法において、最初に行われる独立成分分離処理は、非ガウス性のデータを独立成分に分離するための非ガウス性独立成分分離処理であることを特徴とする音声分離方法である。非ガウス性独立成分分離処理は他の独立成分分処理方法に比べてその前処理としての無相関化処理の影響を受けやすいから、このような第４の発明によれば、最初に非ガウス性独立成分分離処理を行うことにより、無相関化処理がうまく実行されたかどうかを、該無相関化処理に引き続く非ガウス性独立成分分離処理によって効果的に評価することが可能となる。

【００１７】また、本出願に係る第５の発明は、第１乃至第４の発明である音声分離方法において、前記無相関化処理は、少なくとも主成分分析及び因子分析を行うことを特徴とする音声分離方法である。このような第５の

発明によれば、各主成分の寄与率を求めて累積寄与率が所定のしきい値を越えるところの成分数を次数とすることなどにより、採用する主成分データの数（次数）を決定した上で、効果的に無相関化処理を行うことが可能となる。

【００１８】また、本出願に係る第６の発明は、複数発言者の音声データが混在している混在音声データを、発言者毎の音声データに分離し、該発言者毎の音声データにつき発言者を特定する発言者特定方法において、（１）第１乃第５のいずれかの発明の音声分離方法により、複数発言者の音声データが混在している混在音声データを、発言者毎の音声データに分離するステップと、（２）発言者毎に該発言者を特定するための特定パラメータを準備するステップと、（３）分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して、発言者を特定するステップとを有することを特徴とする発言者特定方法である。このような第６の発明によれば、例えば、会議の録音データなどに記録された、複数発言者の音声や雑音などが含まれたの混在音声データを音源ごとに分離し、各分離されたの音声データの発言者を特定することによって、例えば、自動的に会議記録データの作成などを行うことができる。

【００１９】また、本出願に係る第７の発明は、第６の発明である発言者特定方法において、前記特定パラメータは、発言者が母音を発音した際のホルマント周波数であり、分離された前記発言者毎の音声データにつき、ホルマント周波数を求め、求められたホルマント周波数に関して、前記特定パラメータとしてのホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定方法である。このような第７の発明によれば、フーリエ変換などの容易な処理で抽出できる特徴量であるホルマント周波数を用いて、各分離された音声データの発言者特定を容易に行うことができる。

【００２０】また、本出願に係る第８の発明は、第７の発明である発言者特定方法において、前記特定パラメータは、発言者が母音を発音した際の第１ホルマント周波数及び第２ホルマント周波数であり、分離された前記発言者毎の音声データにつき、第１ホルマント周波数及び第２ホルマント周波数を求め、求められた第１ホルマント周波数及び第２ホルマント周波数に関して、前記特定パラメータとしての第１ホルマント周波数及び第２ホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定方法である。このような第８の発明によれば、第１と第２のスペクトルピークである２つのホルマント周波数を用いて発言者の特定を行うことによって、容易に、かつより正確に特定を行うことができる。

【００２１】また、本出願に係る第９の発明は、第６の発明乃至第８の発明のいずれかに記載の発言者特定方法において、分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して発言者を特定するステップにて発言者を特定できなかった場合には、該音声データから複数の時点のホルマント周波数を求め、求められた複数時点のホルマント周波数に関して、前記特定パラメータとしての複数時点のホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定方法である。このような第９の発明によれば、ある音声の発声者を特定する上での特徴量であるホルマント周波数の、時間的変動をも考慮することにより、より正確に発言者の特定を行うことができる。

【００２２】また、本出願に係る第１０の発明は、第６の発明乃至第９の発明のいずれかに記載の発言者特定方法において、分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して発言者を特定するステップにて発言者を特定できなかった場合には、該音声データから有声音データを分離し、該有声音データにつき、ホルマント周波数を求め、求められたホルマント周波数に関して、前記特定パラメータとしてのホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定方法である。ホルマント周波数による発言者の特定は、有声音、特に母音の識別に有効であるので、このような第１０の発明によれば、無声音を含む様々な音声をもより正確に識別することができる。ここで、無相関化処理及び独立成分分離処理がなされる前の音声データが複数人の音声が混在しているデータであるのに対して、無相関化処理及び独立成分分離処理という二つの処理によって分離された分離音声データは、ある一人の音声が抽出されたデータとなっている。よって、このような二つの処理によって分離された分離音声データからは有声音を高い精度で抽出することができる。

【００２３】また、本出願に係る第１１の発明は、第１０の発明の発言者特定方法において、分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して発言者を特定するステップにて発言者を特定できなかった場合には、該音声データから有声音データを分離し、該有声音データにつき、第１ホルマント周波数及び第２ホルマント周波数を求め、求められた第１ホルマント周波数及び第２ホルマント周波数に関して、前記特定パラメータとしての第１ホルマント周波数及び第２ホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定方法である。このような第１１の発明によれば、分離された有声音データに対して、第１と第２のスペクトルピークである２つのホルマント周波数を用いて発言者の特定を行うことによって、より正確に特定を行うことができる。

【００２４】また、本出願に係る第１２の発明は、第１０の発明または第１１の発明の発言者特定方法において、分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して発言者を特定するステップにて発言者を特定できなかった場合には、該有声音データにつき、複数の時点のホルマント周波数を求め、求めら

れた複数時点のホルマント周波数に関して、前記特定パラメータとしての複数時点のホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定方法である。このような第１２の発明によれば、分離された有声音データに対して、発言者特定上の特徴量であるホルマント周波数の時間的変動をも考慮することにより、より正確に発言者の特定を行うことができる。

【００２５】また、本出願に係る第１３の発明は、第１０の発明又は第１１の発明の発言者特定方法において、前記音声データから前記有声音データを分離する際に、該音声データに対して独立成分に分離するための独立成分分離処理が行われることを特徴とする発言者特定方法である。有声音は声帯の振動を伴うものなので、このような第１３の発明によれば、音声データに独立成分分離処理をかけることによって、声帯の振動を伴わない無声音と声帯の振動を伴う有声音とを容易に分離することが可能となる。

【００２６】また、本出願に係る第１４の発明は、複数発言者の音声データが混在している混在音声データから、議事録を作成する議事録作成方法において、第６の発明乃至第１３のいずれかの発明の発言者特定方法により、分離された前記発言者毎の音声データにつき、発言者を特定するステップと、特定された発言者と、該発言者の発言とを対応付けて記録媒体に出力することにより、議事録を作成するステップとを有することを特徴とする議事録作成方法である。このような第１４の発明によれば、発言者の特定が自動的に正確に行われるため、長時間にわたる会議の議事録作成を自動的に行うことができ便利である。

【００２７】また、本出願に係る第１５の発明は、複数発言者の音声データが混在している混在音声データを、発言者毎の音声データに分離する音声データ分離装置において、前記混在音声データを互いに無相関化するために無相関化処理を行い、前記無相関化処理の行われたデータを独立成分に分離するために独立成分分離処理を行うことを特徴とする音声分離装置である。このような第１５の発明によれば、入力される混在音声データ（生データ）に含まれる各音声データの相関性、及び独立性の両性質をともに考慮し、複数の音声データや混入する雑音などの有する相関性や独立性が、時間的・空間的に変動する場合でも、発言者毎の音声に正確に分離することが可能な音声分離装置を実現できるまた、本出願に係る第１６の発明は、第１５の発明である音声分離装置において、前記独立成分分離の行われたデータの分離性が不十分な場合には、分離性が十分になるまで、前記独立成分分離処理の行われたデータについて、前記無相関化処理及び前記独立成分分離処理を繰り返し行うことを特徴とする音声分離装置である。このような第１６の発明によれば、混在音声データを音源毎の音声データに充分に分離させることの可能な音声分離装置を実現できる。

【００２８】また、本出願に係る第１７の発明は、第１５又は第１６の発明である音声分離装置において、データの性質により、前記独立成分分離処理として、非ガウス性のデータを独立成分に分離するための非ガウス性独立成分分離処理、非定常性のデータを独立成分に分離するための非定常性独立成分分離処理、有色性のデータを独立成分に分離するための有色性独立成分分離処理、のうちのいずれかの処理を行うことを特徴とする音声分離装置である。このような第１７の発明によれば、無相関化処理の行われたデータの性質に応じて最適な独立成分分離処理を行うことができるから、混在音声データを音源毎の音声データにより効果的に分離させることの可能な音声分離装置を実現できる。

【００２９】また、本出願に係る第１８の発明は、第１７の発明である音声分離装置において、最初に行われる独立成分分離処理は、非ガウス性のデータを独立成分に分離するための非ガウス性独立成分分離処理であることを特徴とする音声分離装置である。非ガウス性独立成分分離処理は他の独立成分分処理方法に比べてその前処理としての無相関化処理の影響を受けやすいから、このような第１８の発明によれば、最初に非ガウス性独立成分分離処理を行うことにより、無相関化処理がうまく実行されたかどうかを、該無相関化処理に引き続く非ガウス性独立成分分離処理によって効果的に評価することが可能な音声分離装置を実現できる。

【００３０】また、本出願に係る第１９の発明は、第１５乃至第１８の発明である音声分離装置において、前記無相関化処理は、少なくとも主成分分析及び因子分析を行うことを特徴とする音声分離装置である。このような第１９の発明によれば、各主成分の寄与率を求めて累積寄与率が所定のしきい値を越えるところの成分数を次数とすることなどにより、採用する主成分データの数（次数）を決定した上で、効果的に無相関化処理を行うことが可能な音声分離装置を実現できる。

【００３１】また、本出願に係る第２０の発明は、複数発言者の音声データが混在している混在音声データを、発言者毎の音声データに分離し、該発言者毎の音声データにつき発言者を特定する発言者特定装置において、第１５乃至第１９のいずれかの発明の音声分離装置により、複数発言者の音声データが混在している混在音声データを、発言者毎の音声データに分離し、分離された前記発言者毎の音声データにつき、発言者毎に該発言者を特定するための特定パラメータを参照して発言者を特定することを特徴とする発言者特定装置である。このような第２０の発明によれば、例えば、会議の録音データなどに記録された、複数発言者の音声や雑音などが含まれたの混在音声データを音源ごとに分離し、各分離されたの音声データの発言者を特定することによって、例えば、自動的に会議記録データの作成などを行うことの可能な発言者特定装置が実現できる。

【００３２】また、本出願に係る第２１の発明は、第２０の発明である発言者特定装置において、前記特定パラメータは、発言者が母音を発音した際のホルマント周波数であり、分離された前記発言者毎の音声データにつき、ホルマント周波数を求め、求められたホルマント周波数に関して、前記特定パラメータとしてのホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定装置である。このような第２１の発明によれば、フーリエ変換などの容易な処理で抽出できる特徴量であるホルマント周波数を用いて、各分離された音声データの発言者特定を容易に行うことの可能な発言者特定装置が実現できる。

【００３３】また、本出願に係る第２２の発明は、第２１の発明である発言者特定装置において、前記特定パラメータは、発言者が母音を発音した際の第１ホルマント周波数及び第２ホルマント周波数であり、分離された前記発言者毎の音声データにつき、第１ホルマント周波数及び第２ホルマント周波数を求め、求められた第１ホルマント周波数及び第２ホルマント周波数に関して、前記特定パラメータとしての第１ホルマント周波数及び第２ホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定装置である。このような第２２の発明によれば、第１と第２のスペクトルピークである２つのホルマント周波数を用いて発言者の特定を行うことによって、容易に、かつより正確に特定を行うことの可能な発言者特定装置が実現できる。

【００３４】また、本出願に係る第２３の発明は、第２０の発明乃至第２２の発明のいずれかに記載の発言者特定装置において、分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して発言者を特定できなかった場合には、該音声データから複数の時点のホルマント周波数を求め、求められた複数時点のホルマント周波数に関して、前記特定パラメータとしての複数時点のホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定装置である。このような第２３の発明によれば、ある音声の発声者を特定する上での特徴量であるホルマント周波数の、時間的変動をも考慮することにより、より正確に発言者の特定を行うことの可能な発言者特定装置が実現できる。

【００３５】また、本出願に係る第２４の発明は、第２０の発明乃至第２３の発明のいずれかに記載の発言者特定装置において、分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して発言者を特定できなかった場合には、該音声データから有声音データを分離し、該有声音データにつき、ホルマント周波数を求め、求められたホルマント周波数に関して、前記特定パラメータとしてのホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定装置である。ホルマント周波数による発言者の特定は、有声音、特に母音の識別に有効であるので、このような第２４の発明によれば、無声音を含む様々な音声をもより正確に識別することができる。ここで、無相関化処理及び独立成分分離処理がなされる前の音声データが複数人の音声が混在しているデータであるのに対して、無相関化処理及び独立成分分離処理という二つの処理によって分離された分離音声データは、ある一人の音声が抽出されたデータとなっている。よって、このような二つの処理によって分離された分離音声データからは有声音を高い精度で抽出することができる。

【００３６】また、本出願に係る第２５の発明は、第２４の発明の発言者特定装置において、分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して発言者を特定できなかった場合には、該音声データから有声音データを分離し、該有声音データにつき、第１ホルマント周波数及び第２ホルマント周波数を求め、求められた第１ホルマント周波数及び第２ホルマント周波数に関して、前記特定パラメータとしての第１ホルマント周波数及び第２ホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定装置である。このような第２５の発明によれば、分離された有声音データに対して、第１と第２のスペクトルピークである２つのホルマント周波数を用いて発言者の特定を行うことによって、より正確に特定を行うことの可能な発言者特定装置が実現できる。

【００３７】また、本出願に係る第２６の発明は、第２４の発明または第２５の発明の発言者特定装置において、分離された前記発言者毎の音声データにつき、前記特定パラメータを参照して発言者を特定できなかった場合には、該有声音データにつき、複数の時点のホルマント周波数を求め、求められた複数時点のホルマント周波数に関して、前記特定パラメータとしての複数時点のホルマント周波数を参照して、発言者を特定することを特徴とする発言者特定装置である。このような第２６の発明によれば、分離された有声音データに対して、発言者特定上の特徴量であるホルマント周波数の時間的変動をも考慮することにより、より正確に発言者の特定を行うことの可能な発言者特定装置が実現できる。

【００３８】また、本出願に係る第２７の発明は、第２４の発明又は第２５の発明の発言者特定装置において、前記音声データから前記有声音データを分離する際に、該音声データに対して独立成分に分離するための独立成分分離処理が行われることを特徴とする発言者特定装置である。有声音は声帯の振動を伴うものなので、このような第２７の発明によれば、音声データに独立成分分離処理をかけることによって、声帯の振動を伴わない無声音と声帯の振動を伴う有声音とを容易に分離することの可能な発言者特定装置が実現できる。

【００３９】また、本出願に係る第２８の発明は、複数発言者の音声データが混在している混在音声データから、議事録を作成する議事録作成装置において、第２０

乃至第２７のいずれかの発明の発言者特定装置により、分離された前記発言者毎の音声データにつき、発言者を特定し、特定された発言者と、該発言者の発言とを対応付けて記録媒体に出力することにより、議事録を作成することを特徴とする議事録作成装置である。このような第２８の発明によれば、発言者の特定が自動的に正確に行われるため、長時間にわたる会議の議事録作成を自動的に行うことの可能な議事録作成装置が実現できる。

【００４０】また、第１乃至第５のいずれかの発明の音声分離方法を音声分離装置に実行させるためのコンピュータプログラムも実現可能である。

【００４１】また、第６乃至第１３のいずれかの発明の発言者特定方法を発言者特定装置に実行させるためのコンピュータプログラムも実現可能である。

【００４２】また、そのようなコンピュータプログラムを記録したコンピュータ読み取り可能な記録媒体も実現可能である。

【００４３】

【発明の実施の形態】＝＝混在音声データの音声分離＝＝

以下、図面を参照しつつ、本発明のより具体的な実施形態につき、詳細に説明する。まず、本発明の方法の前半部分である、混在音声データの音声分離ステップについて説明する。

【００４４】本実施形態では、２人で行われたある会議の発言内容の音声データを２本のマイク（マイク１、マイク２）で拾う。図１は、そのうちマイク１から入力された音声データ（生データ）Ｘの波形である。この混在音声データには、複数の発言者の音声データが混在しているのみならず、音楽や、さらには雑音などが混ざっていてもよい。２人の発声をそれぞれ音源Ｓ１、Ｓ２と呼ぶことにする。

【００４５】図２は、音声分離処理のサイクルを示す図である。マイク１及びマイク２から入力された混在音声データは、まず無相関化処理Ｗ１にかけられる。無相関化処理Ｗ１に渡される音声データは、図１の[１]、[２]のようにセグメント化されて１つずつ渡される。最も効率がよいように、各セグメントは互いに１／２周期ずつオーバーラップしている。

【００４６】図２において、無相関化処理Ｗ１の次のステップであるＩＣチューナーは、独立成分解析（ＩＣＡ）の手法を３種類のうちから選択するためのチューナーである。その次のステップである独立成分分離処理Ｗ２は、非ガウス性に基づく分離処理Ｗ２（α）、非定常性に基づく分離処理Ｗ２（β）、有色性に基づく分離処理Ｗ（γ）の３種類のうちいずれかの方式の処理を行う。Ｗ２の後のステップの評価器Ｅでは、Ｗ２にて分離されたデータの分離性の評価を行う。マイクから入力された混在音声データの音声分離性能が充分になるまで、以上のＷ１→ＩＣチューナー→Ｗ２→Ｅというサイクルを繰り返し回す。ただし、１回目のサイクルでは、独立成分分離処理Ｗ２として、非ガウス性に基づく独立成分分離処理Ｗ２（α）を行い、２回目以降のサイクルでは、ＩＣチューナの選択に従って、Ｗ２（α）、Ｗ２（β）、Ｗ２（γ）の３種類のうちから適切な方式の独立成分分離処理を行う。

【００４７】図３は、１回目の音声分離サイクルを示している。図１における前記[１]の時間セグメントの、マイク１及びマイク２からの混在音声データｘ１、ｘ２が、まず無相関化処理Ｗ１に入力される。

【００４８】図７及び図８は、それぞれｘ１及びｘ２のデジタル化波形図データ（縦軸は音の強さで、単位はミリボルト）を示す。各時点のｘ１、ｘ２データを、横軸をｘ１の強さ、縦軸をｘ２の強さとして散布図を描くと図９のようになる。散布図は、第１象限から第３象限にかけて若干直線的な分布を呈し、ｘ１とｘ２のデータは互いに相関性を有することを示している。これら生データであるｘ１、ｘ２が無相関化処理Ｗ１にかけられると、互いに相関性を有しないデータｆ１、ｆ２に変換される。

【００４９】ｆ１及びｆ２の散布図を図１０に示す。図１０の横軸は因子得点Ｆの第１因子ｆ１、縦軸は因子得点Ｆの第２因子ｆ２を示している。図９が軸に対していびつな平行四辺形状に分布していたのに対し、軸に対してまっすぐで形の整ったひし形状に分布しており、ｆ１とｆ２はもはや互いに相関性を有していないことがわかる。

【００５０】ここで、無相関化処理の内容について説明する。図６は、無相関化処理Ｗ１の一例のフローチャートを示したものである。まず、図７及び図８に示した音声生データｘ１、ｘ２を（１）式により標準化する。標準化の結果、平均が０、標準偏差１のデータとなる。

【数１】

$$\mathbb{X} = \frac{(\mathbb{X} - \mu)}{\sigma} \qquad (1)$$

【００５１】生データｘ１、ｘ２の相関行列(ベクトルＣ)を（２）式より求める。（２）式において（ｘ１、ｘ２）はベクトルの内積を表す。

【数２】

$$\mathbb{C} = \frac{1}{n}(\mathbb{X}_1, \mathbb{X}_2) \qquad (2)$$

【００５２】上記相関行列に対する固有値λｉと固有ベクトルＡを（３）より求める。

【００５３】

【数３】

$$\mathbb{C}\mathbb{A} = \lambda\mathbb{A} \qquad (3)$$

【００５４】今、因子分析によって、互いに無相関な因子得点を求めようとしているのだが、その際、第１番目の因子から始めて、何番目の因子までを採用するのかが重要な点である。ｍ番目の因子までを採用する場合を、

m次元と呼ぶ。先に求めた固有ベクトルＡにより、（４）式によって主成分Ｚが求まる。

【数４】

$$\mathbb{Z} = \mathbb{A}\,\mathbb{X} \qquad (4)$$

【００５５】次にm個の因子に対して、（５）式の形の定義式にて因子分析を実行する。（５）式におけるｅは、特殊因子と呼ばれるものである。

【数５】

$$\mathbb{X} = \mathbb{A}\,\mathbb{Z} + \mathbb{e} \qquad (5)$$

【００５６】この因子モデルが（６）式の表現をとる。（６）式における因子負荷量ｂｉｊ、因子得点Ｆは、（７）式及び（８）式によって求める。そして、図６のフローチャートの最終ステップで、結局音声生データは、互いに無相関な因子得点（ベクトルＦ）に変換される。

【数６】

$$\mathbb{X} = \mathbb{b}\,\mathbb{F} + \mathbb{e} \qquad (6)$$

【数７】

$$\mathbb{b}ij = \sqrt{\lambda j}\,\mathbb{A}ij \qquad (7)$$

【数８】

$$\mathbb{F} = (\mathbb{b}'\mathbb{b})^{-1}\mathbb{b}'\mathbb{X} \qquad (8)$$

【００５７】以上説明したＷ１の主な特徴は、主成分分析と因子分析とを組み合わせている点である。その効果は、主成分分析を実行すると各主成分の寄与率を同時に求めることができるので、例えば、第１次主成分から第m次主成分までの累積寄与率が８０％を超えるまでの主成分を採用するようにすることで、次数mを決定することにある。分離すべき音声生データは、時間的変動が大きく、混合による相関の度合いが大きく変化するので、何個の因子を採用するかは無相関化処理において重要な点である。

【００５８】発話者の人数があらかじめ判明している場合には、次数mを発話者の人数に固定してしまえばよいが、人数が不明なときは、例えば、累積寄与率が所定のしきい値を超えたときの主成分数を次数mとする。次数mの決定方法は、システムに応じて様々な方法を準備しておき、臨機応変に変化させる（チューニングする）ことが好ましい。次にこのチューニングの一実施例について詳しく説明する。

【００５９】図２３は、システムに応じた方法で次数mを決定する手順を示すフローチャートである。図２３で、ＲＫ０は累積寄与率の初期しきい値、Ｍは採用し得る最大次数（次数の上側しきい値）、△ＲＫは累積寄与率の変化量である。主成分分析を実行すると、図２１のような、次数m（第m主成分まで採用したということを示す）とその累積寄与率との関係を示すグラフが得られる。図２１にはＡ、Ｂ、Ｃ３種類のグラフの例を描いている。

【００６０】まず、第１の処理ステップとして、累積寄与率ＲＫにしきい値ＲＫ０（この実施例では８０％）を設定しておき、このしきい値ＲＫ０を超える次数mを求める。ところが、次数があまりに大きいとその後の処理が煩雑に過ぎるので、あらかじめ次数の上限値Ｍを決めておく。図２１の例では、Ｍ＝４とすると、Ａの場合はしきい値ＲＫ０を超える次数m＝２であるので、m＝２＜４＝Ｍとなって、次数mは２に決定される。Ｂの例ではＲＫ０を超える次数mは５であるので、m＝５＞４＝Ｍとなってしまい、次数mはまだ決定されない。Ｃの例でも同様に次数mは決定されない。

【００６１】そのような場合は図２２に示す、第２のステップを実行する。すなわち、次数mの増加に対する、ＲＫの差分変化量△ＲＫを調べる。これは要するに、累積寄与率の変化が最大となる次数mをもって採用すべき次数とするという処理方法である。この実施例では、Ｂの例ではm＝２、Ｃの例ではm＝４において△ＲＫが最大値をとる。この場合も次数mが上限値Ｍよりも下ならば、その次数mを採用とするが、Ｍを上回る場合は、その処理が次のステップに送られる。

【００６２】第２のステップでも次数mが上限値Ｍを超えてしまう場合であれば、次に累積寄与率のしきい値ＲＫ０を引き下げて、例えば６０％（＝ＲＫ１）とし、上記第１のステップと同じように比較する。新しいしきい値ＲＫ１を超えるところの次数がＭ＝４以下であれば、これを次数mとして採用とし、Ｍを超える場合は、所定の下げ幅で順次ＲＫ２、ＲＫ３、・・・ＲＫｎの値を下げる。ただし、累積寄与率ＲＫが５０％を下回るということは、半分以上の情報が失われてしまうことを意味するので、ＲＫｎの下限値は５０％とする。

【００６３】次数mがＲＫｎ＝５０％以上で、かつＭ以下の値で発見されない場合は、再び上記第２のステップと同様の処理、すなわち△ＲＫが最大になる次数を求めて、その値を次数mとして採用してしまう。これは、累積寄与率が大きく変化するということは、その次数の前後で情報がより多く保存されるということを意味するので、少なくともその次数までは採用したい、という考えに基づくものである。

【００６４】以上のようにして、図３において、無相関化されたデータｆ１、ｆ２は、ただちに独立成分分離処理Ｗ２に送られる。１回目の音声分離サイクルでは、これらの無相関化データｆ１、ｆ２に対し、非ガウス性に基づく独立成分分離処理Ｗ２（α）を実行する。

【００６５】以上、図３におけるＷ１及びＷ２(α)の処理により、分離信号ａおよびｂが得られ、これらの分離性（充分に分離されているか否か）を評価器Ｅで評価し、分離が不十分なとき（図の＊１）はこれらａ、ｂのデータに対して、２回目のサイクルを実行する。

【００６６】２回目のサイクルの例を図４に示す。図３に示した１回目のサイクルと似ているが、ＩＣチューナーにおける処理が加わっている。独立成分分離処理Ｗ２

を行う前に、ＩＣチューナーで２回目の無相関化処理されたデータｆ１´，ｆ２´の信号特性を解析し、非ガウス性に基づく処理Ｗ２(α)、非定常性に基づく処理Ｗ２(β)、有色性に基づく処理Ｗ２（γ）のいずれをＷ２として実行するかを選択する。この例ではＷ２（β）を実

された入力データのいずれもが有色性を有すると評価された場合、ＩＣチューナーは、独立成分分離処Ｗ２として有色性に基づく処理Ｗ２(γ)を選択する。

【００７１】図５は３回目のサイクルを示している。各処理は２回目のサイクルと同様であるが、３回目の独立